

Satellite Imagery-Based Property Price Prediction Using Multimodal Learning

A Multimodal Regression Approach Using Tabular
Housing Data and Satellite Images

Project Report

Submitted By:

Gopal Datt Sharma
23115047
Electrical (III year)

1. Overview

Accurate real estate valuation depends not only on property-specific attributes such as size, number of rooms, and construction quality, but also on the surrounding environmental and neighborhood context. Traditional pricing approaches based solely on tabular data often fail to capture visual cues such as green cover, proximity to water, road density, and urban layout, all of which significantly influence market value.

To address this limitation, this project develops a **multimodal regression framework** that integrates **tabular housing data with satellite imagery** for residential property price prediction. Latitude and longitude information is used to programmatically fetch satellite images for each property, enabling the incorporation of visual environmental context alongside structured numerical data.

The modeling pipeline begins with a **tabular-only baseline**, where a **Random Forest regressor** is trained using core housing attributes such as size, room configuration, quality indicators, and geographic coordinates. To enhance the expressive power of the tabular data, extensive **feature engineering** is performed based on insights from exploratory and geospatial analysis. These engineered features capture layout efficiency (e.g., space utilization ratios), neighborhood context (relative size compared to nearby properties), temporal effects (house age and renovation history), functional room configuration, quality-size interactions, and geographic influence (distance to city center and spatial interactions). Log transformations are applied to skewed variables to stabilize learning and reduce the impact of extreme values.

For the multimodal model, **satellite images are programmatically downloaded using geographic coordinates** and processed through a **Convolutional Neural Network (CNN)** to extract high-level visual embeddings representing environmental context such as green cover, road density, water proximity, and neighborhood structure. These image-derived features are then **fused with the engineered tabular features using a feature-level (early) fusion strategy**, allowing the model to jointly learn from visual and numerical information.

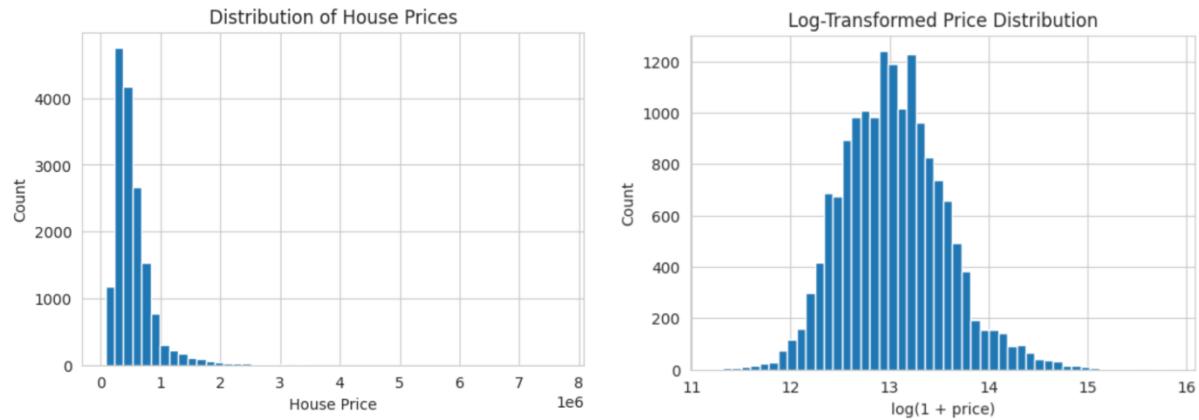
A **CatBoost regressor** is trained on the combined feature space to produce final price predictions. Model performance is evaluated by comparing the tabular-only baseline against the multimodal approach, demonstrating the added value of incorporating satellite imagery.

To enhance interpretability, **Grad-CAM-based visual explanations** are generated to highlight regions of satellite imagery that contribute most to the predicted property prices. Overall, this project demonstrates how integrating satellite imagery with traditional tabular data can improve real estate valuation by jointly modeling **structural attributes** and **environmental characteristics** within a unified predictive framework.

2. EDA

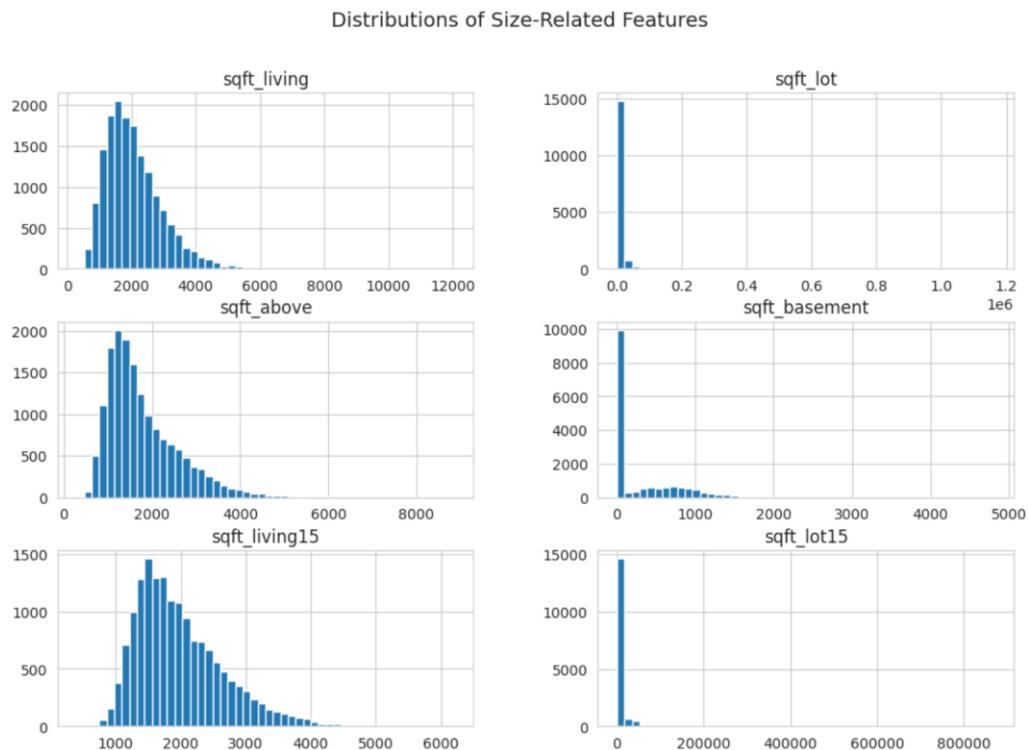
2.1 TABULAR EDA

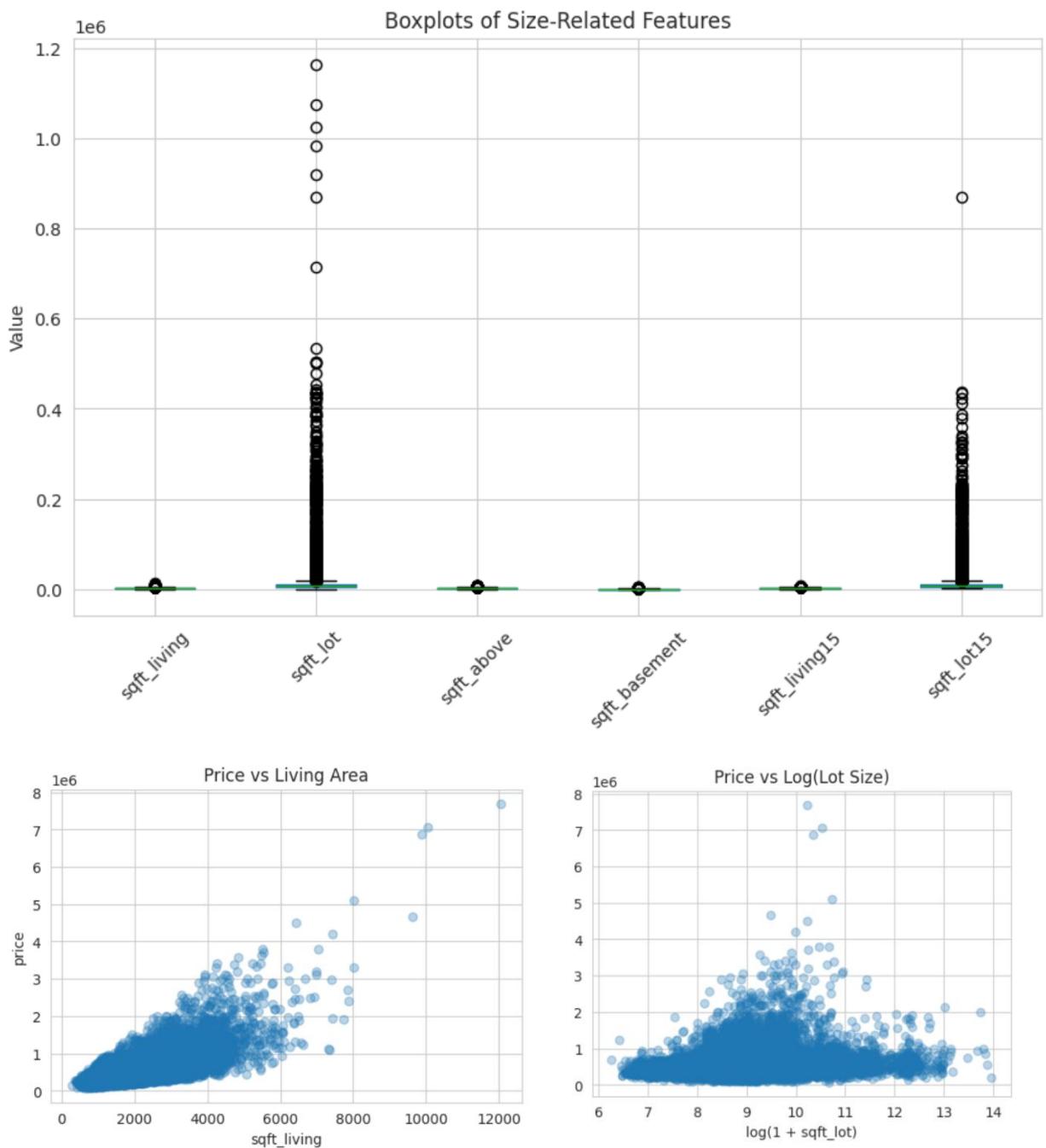
Price Distribution



House prices show a strong right-skew, with most properties concentrated in the lower price range and a small number of high-value outliers. Applying a log transformation produces a more symmetric distribution, making price behavior easier to analyze and model.

Size & Structural Features

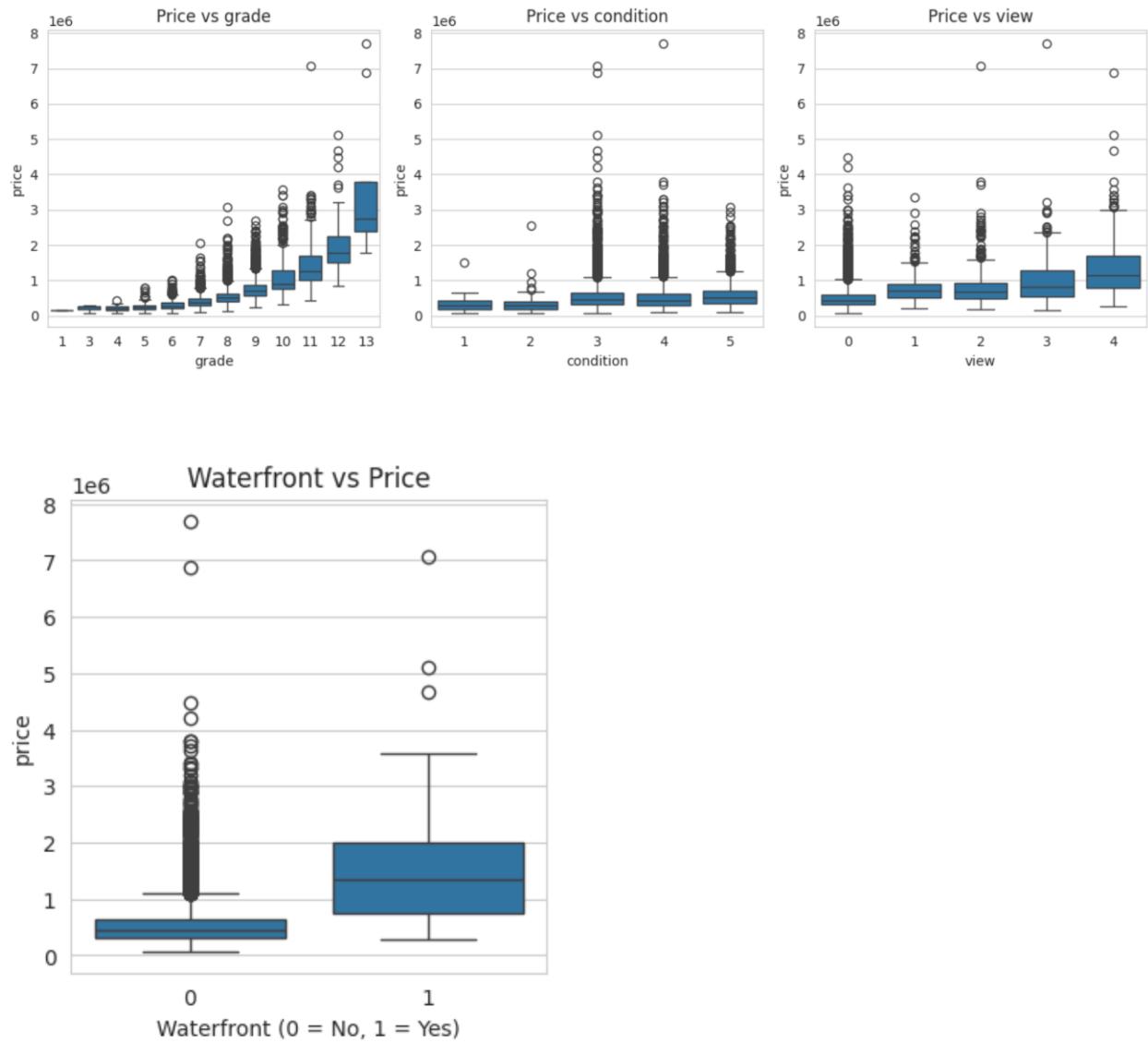




The distributions of size-related features such as living area, lot size, and basement area are right-skewed, indicating that most properties fall within moderate size ranges with a small number of very large homes and lots. Boxplots further highlight the presence of extreme outliers, especially for lot-related features.

The scatter plot between price and living area shows a clear positive relationship, confirming that larger living spaces are generally associated with higher property prices. In contrast, the relationship between price and lot size (after log transformation) is weaker and more dispersed, suggesting that land size alone is a less consistent driver of price compared to built-up living area.

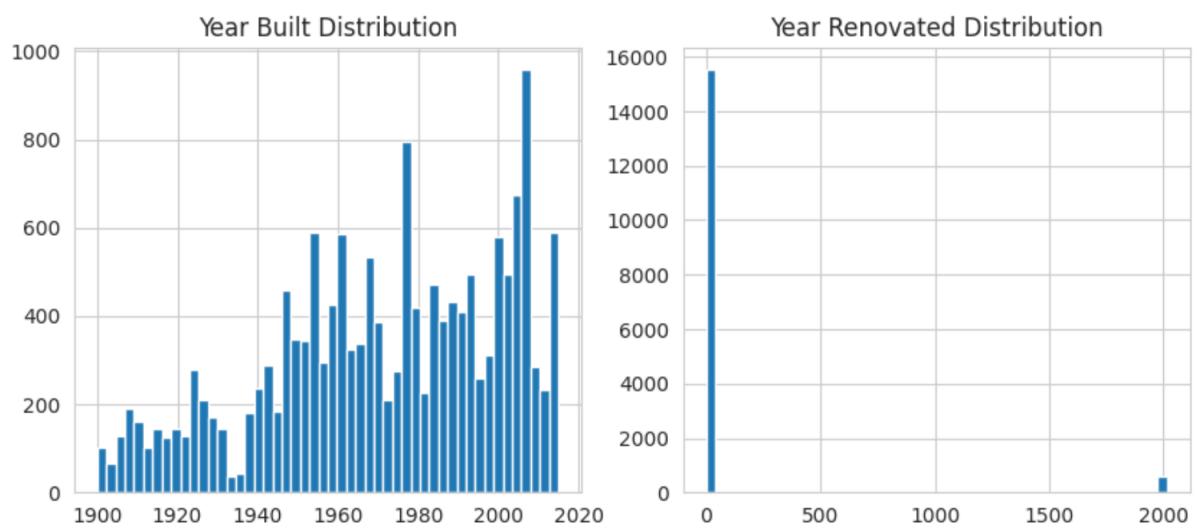
Quality & Categorical Effects



Properties with higher construction grade show a clear and consistent increase in price, indicating that build quality is a strong determinant of property value. While condition also influences price, its effect is comparatively weaker and exhibits greater overlap across categories.

Houses with better views and waterfront access command noticeably higher prices, with waterfront properties showing a distinct upward shift in price distribution. These patterns highlight the importance of qualitative and location-related attributes in real estate valuation.

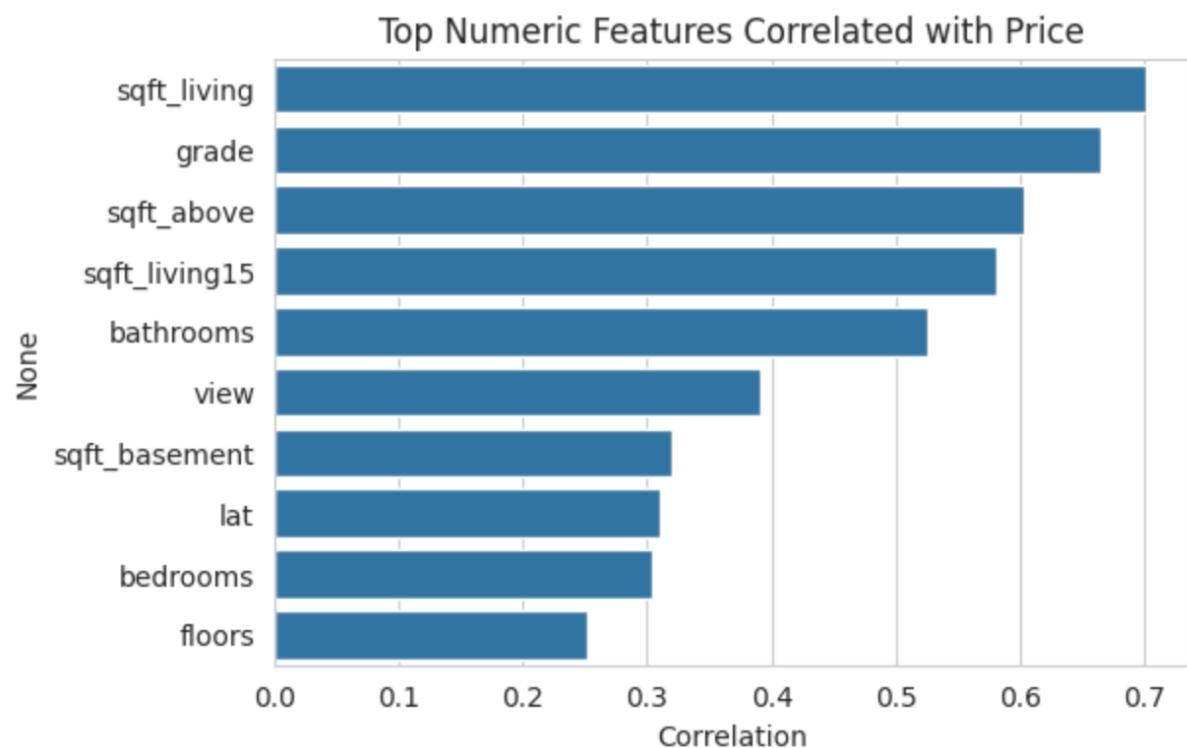
Temporal Features



The year-built distribution shows that a large proportion of properties were constructed in the mid to late 20th century, with fewer homes built in earlier periods. This indicates a mix of older and relatively modern housing stock within the dataset.

The year-renovated distribution reveals that most properties have not been renovated, while a smaller subset has undergone renovations in more recent years. This suggests that renovation status is a distinguishing factor for certain properties and may contribute to price variation.

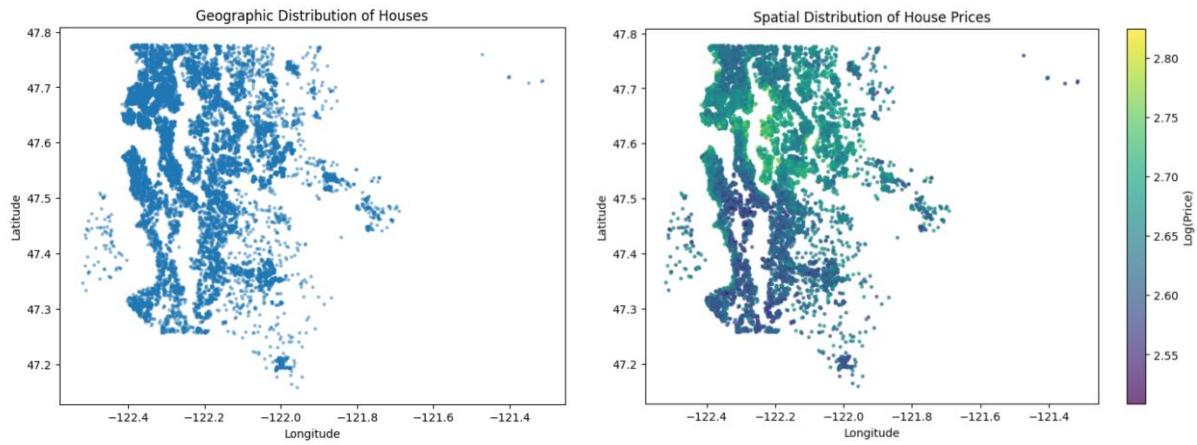
Correlation Summary



Living area and construction grade exhibit the strongest positive correlation with house prices, indicating that built-up space and build quality are key drivers of value. Other features such as bathrooms, view, and basement area show moderate correlations, while location-related attributes also contribute to price variation.

2.2 GEOSPATIAL EDA

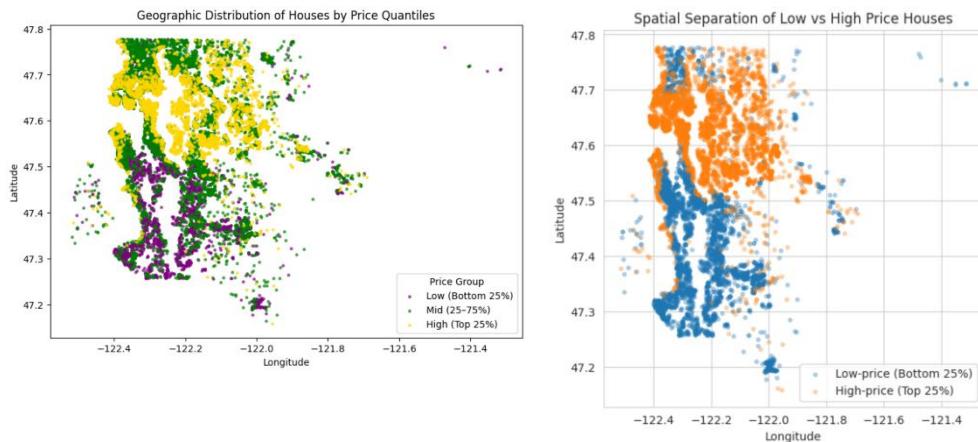
Spatial Distribution



The spatial distribution of properties shows clear geographic clustering, indicating that housing locations are not uniformly distributed across the region. Dense clusters appear around urban and suburban areas, reflecting patterns of residential development.

When prices are overlaid spatially, higher-priced properties tend to concentrate in specific geographic zones, while lower-priced homes are more dispersed. This highlights the strong influence of location on property value and motivates the inclusion of spatial and visual context in the modeling process.

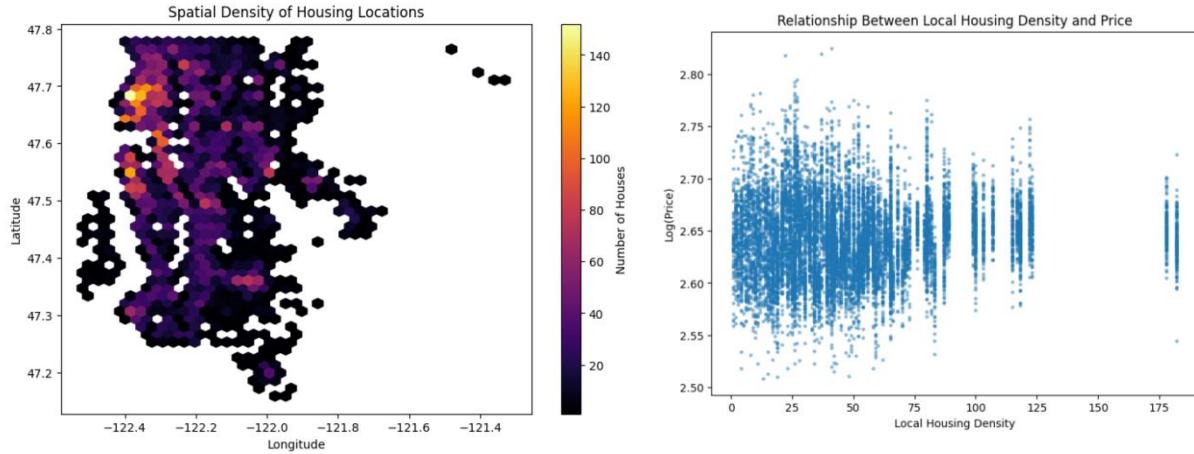
Price Stratification



The geographic distribution of houses by price quantiles reveals clear spatial separation between low-, mid-, and high-priced properties. Higher-priced homes tend to cluster in specific regions, while lower-priced properties are more prevalent in distinct geographic zones.

The direct comparison between low-priced and high-priced houses further emphasizes this spatial separation, indicating that location plays a critical role in price differentiation. These patterns suggest that neighborhood-level characteristics strongly influence property values.

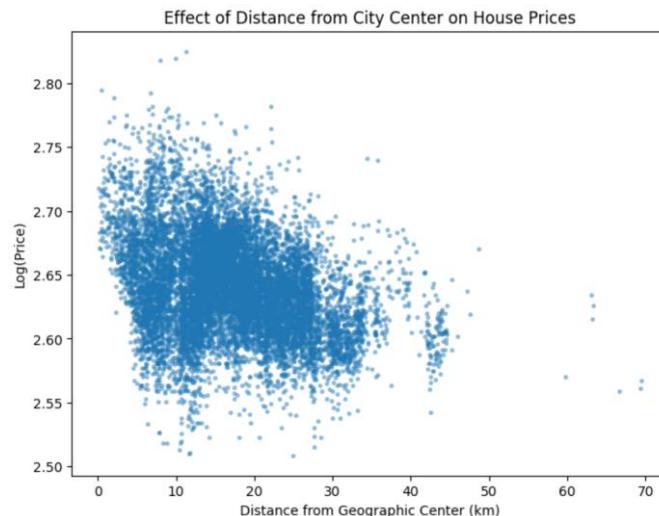
Density & Urban Effects



The spatial density map highlights regions with a high concentration of residential properties, primarily corresponding to urban and suburban areas. Less dense regions appear more scattered, reflecting lower housing concentration.

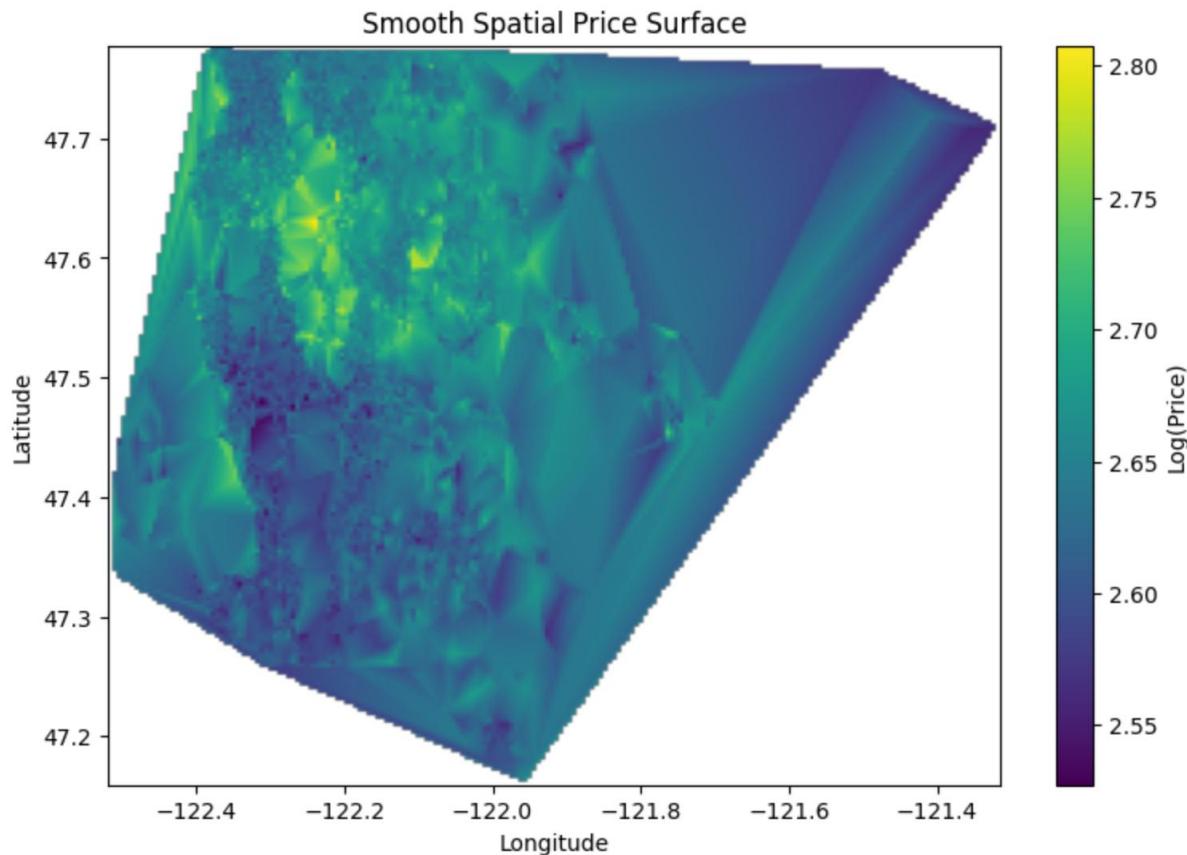
The relationship between local housing density and price shows a broad spread with no strictly linear trend, indicating that higher density alone does not guarantee higher prices. This suggests that while density reflects urban development, additional neighborhood and environmental factors influence property value.

Distance-Based Effect



Property prices generally decrease as distance from the city center increases, indicating a negative relationship between centrality and value. While there is considerable variation, higher-priced properties are more concentrated closer to the urban core, highlighting the importance of accessibility and location in real estate pricing.

Smoothed Spatial Trend



The smoothed spatial price surface reveals clear regional patterns in property values, with higher-priced zones forming contiguous geographic regions rather than isolated points. These smooth gradients indicate that housing prices change gradually across space, reflecting neighborhood-level effects rather than random variation.

This visualization reinforces the importance of spatial context in real estate valuation and supports the inclusion of location-aware and visual features in the modeling framework.

Randomly Sampled Satellite Images



Random satellite image samples show substantial visual diversity across properties, including variations in surrounding greenery, road layouts, building density, and neighborhood structure. These differences indicate that satellite imagery captures contextual and environmental information not present in tabular features alone.

This motivates the use of a CNN to extract visual features that complement traditional housing attributes.

3. Financial / Visual Insights

3.1 Environmental Context from Satellite Imagery

Randomly Sampled Satellite Images from the Dataset



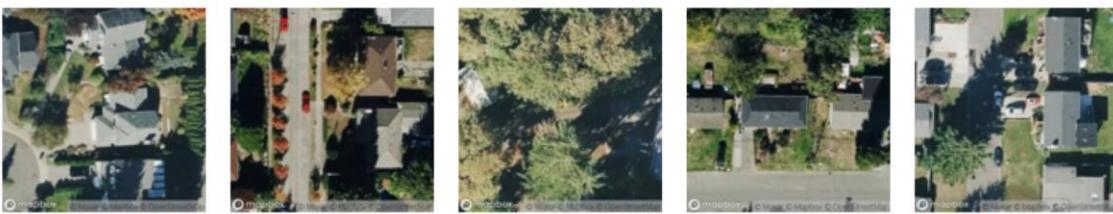
The randomly sampled satellite images highlight substantial visual diversity across properties, including variations in green cover, road layout, housing density, and surrounding open spaces. Some locations exhibit dense vegetation and spacious plots, while others show compact housing with limited greenery and higher road congestion. These visual patterns capture neighborhood characteristics and environmental context that are not fully represented by tabular features alone, motivating the use of satellite imagery as an additional information source in property valuation.

3.2 Visual Impact of Waterfront Proximity

Waterfront Properties (Satellite Images)



Non-Waterfront Properties (Satellite Images)



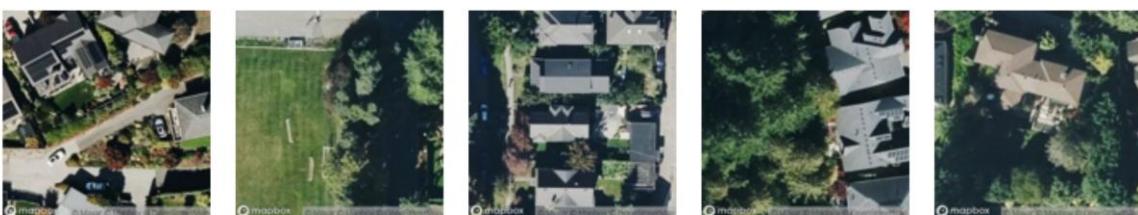
Satellite images of waterfront properties clearly show direct adjacency to water bodies, private docks, open water views, and lower visual obstruction from neighboring structures. In contrast, non-waterfront properties are typically surrounded by dense housing, road networks, and greater structural congestion. These visual differences highlight how proximity to water introduces strong environmental and aesthetic cues that are difficult to capture through tabular features alone, helping explain the consistent price premium associated with waterfront locations.

3.3 Visual Differences Between Low- and High-Priced Properties

Low-Priced Properties (Bottom 25%)



High-Priced Properties (Top 25%)



Satellite images reveal clear visual contrasts between low- and high-priced properties. Lower-priced homes are typically located in denser layouts with smaller plots, limited greenery, and closer proximity to neighboring structures or roads. In contrast, higher-priced properties often exhibit larger lot sizes, more open green spaces, mature tree cover, and lower surrounding density. These visual patterns indicate that environmental openness, vegetation, and spatial isolation are strong contributors to property value and are effectively captured through satellite imagery.

3.4 Grad-CAM-Based Visual Explainability Analysis

To understand **what visual cues the CNN learns from satellite imagery**, Grad-CAM was applied to representative samples from the **bottom 10% (low-priced)** and **top 10% (high-priced)** properties. This highlights the spatial regions that most influenced the model's price predictions.

Low-Priced Properties (Bottom 10%)

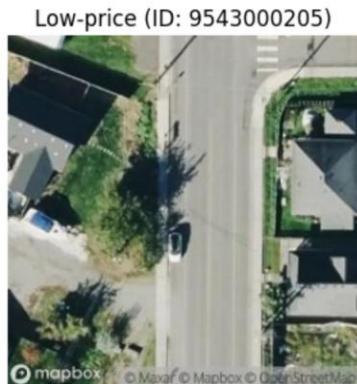
The Grad-CAM visualizations for low-priced houses show attention concentrated on:

- **Road intersections and nearby traffic corridors**
- **Dense surrounding structures and limited open space**
- **Shadows and irregular land usage patterns**

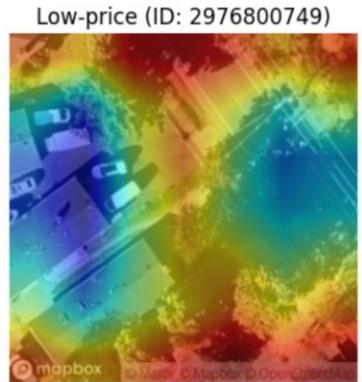
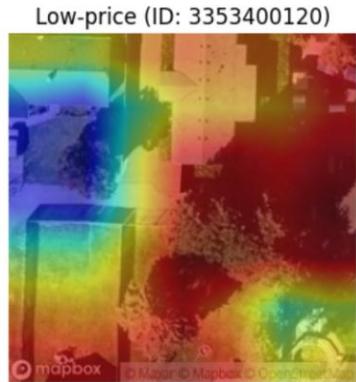
In these images, the model focuses less on the house structure itself and more on **contextual negatives**, such as proximity to roads, congestion, and constrained surroundings. Green areas and open plots receive comparatively lower activation, suggesting limited positive environmental contribution.

This indicates that the CNN learns to associate **high built density, road exposure, and lack of greenery** with lower property valuations.

Low-Price Houses (Bottom 10%)



Grad-CAM — Low-Price Houses



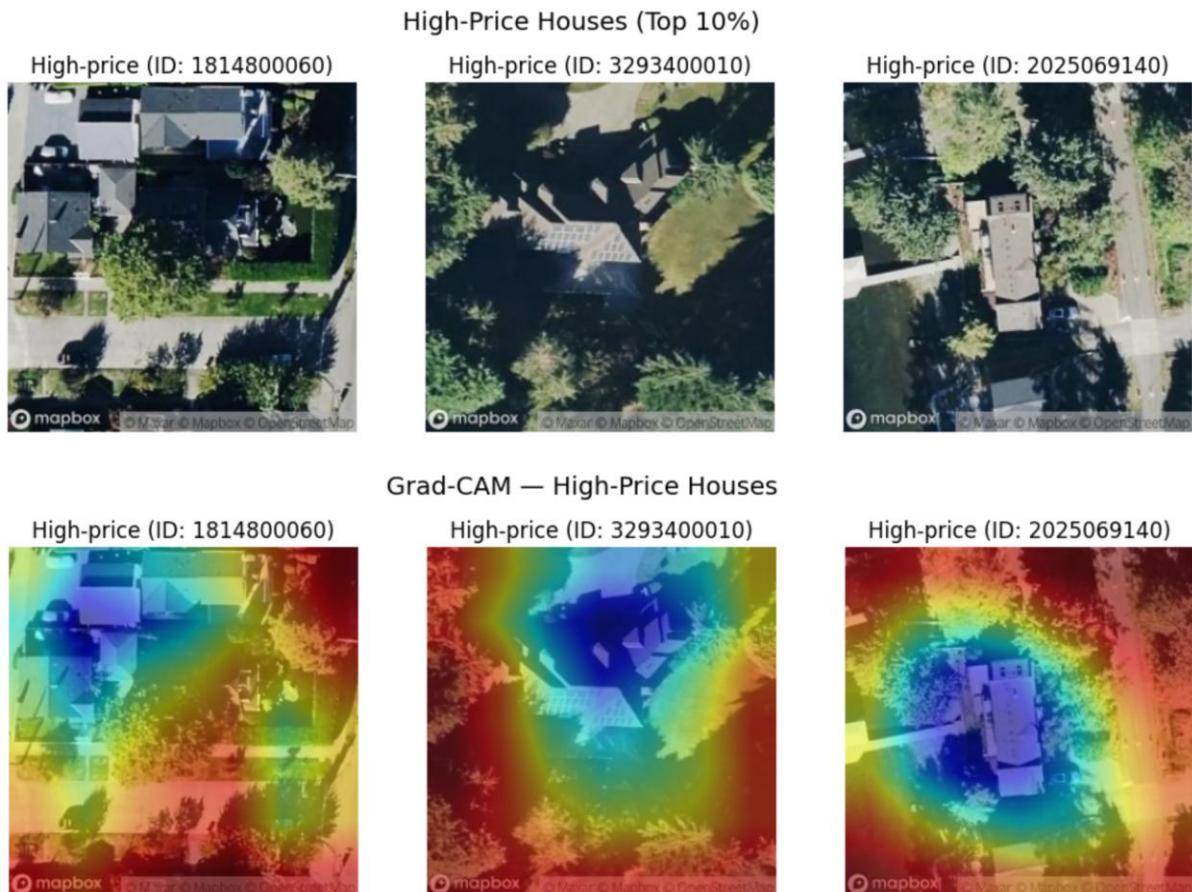
High-Priced Properties (Top 10%)

For high-priced houses, Grad-CAM highlights a different pattern:

- **Large roof footprints and clear property boundaries**
- **Extensive surrounding greenery and tree cover**
- **Open spaces, lawns, and low neighboring density**

The strongest activations appear around the **house core and landscaped surroundings**, rather than roads or neighboring structures. In some cases, the attention spreads symmetrically around the property, capturing **spatial openness and environmental quality**.

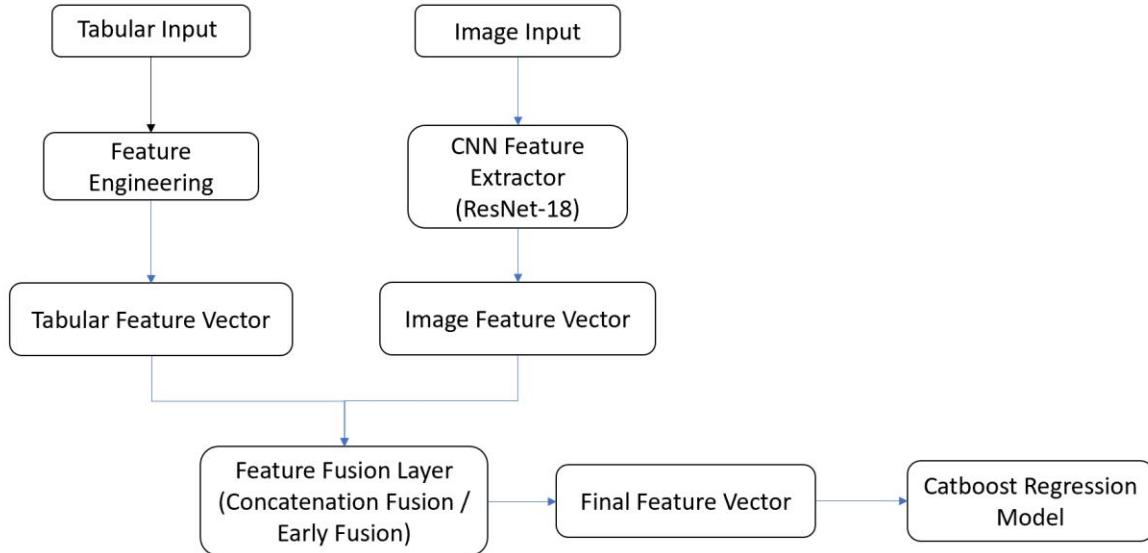
This suggests that the model associates **green cover, privacy, and spacious layouts** with higher property prices.



Key Visual Insights

- The CNN does **not simply detect the house**, but learns **neighborhood-level visual context**.
- **Greenery vs. concrete, open space vs. congestion, and privacy vs. exposure** emerge as strong visual drivers of value.
- These patterns align with domain intuition and validate the usefulness of satellite imagery beyond tabular location features.

4. Architecture Diagram



The proposed system follows a **multimodal early-fusion architecture** that combines tabular housing attributes with visual features extracted from satellite imagery.

Tabular input features are first processed through a **feature engineering pipeline**, where domain-informed transformations and interaction features are created to produce a structured **tabular feature vector**.

In parallel, satellite images corresponding to each property are passed through a **ResNet-18 CNN**, used as a fixed feature extractor to obtain high-level **image feature embeddings** capturing visual context such as greenery, road structure, and surrounding layout.

The tabular and image feature vectors are then **concatenated using an early fusion strategy**, forming a unified representation that captures both structural and visual information.

This fused feature vector is finally fed into a **CatBoost regression model**, which learns non-linear relationships across modalities and outputs the predicted house price.

This architecture enables effective integration of numerical and visual cues while maintaining interpretability and modularity.

5. Results

Evaluation Setup

To evaluate the contribution of satellite imagery, two models were compared using the same train–test split and target variable. Performance was measured using RMSE on price and R² score on log-transformed price to account for skewness in housing prices.

Model Comparison

- Tabular-only baseline: Random Forest trained on engineered tabular features
- Multimodal model: Satellite image features extracted using ResNet-18 and fused with engineered tabular features, followed by CatBoost regression

Model	RMSE (Price) ↓	R ² (log price) ↑
Tabular Only (Random Forest)	139,195.04	0.8751
Tabular + Satellite Images (ResNet-18 + CatBoost)	114,039.08	0.8987

Result Interpretation

The multimodal model significantly outperforms the tabular-only baseline, achieving a **reduction of ~25,000 in RMSE** and an **increase of 0.0236 in R² score**. This improvement demonstrates that satellite imagery provides complementary information beyond traditional structural and geographic features.

Role of Visual Information

The performance gain is supported by visual and explainability analysis. Satellite images capture high-level contextual cues such as:

- Green cover and tree density
- Plot openness and surrounding built density
- Waterfront proximity and road adjacency

Grad-CAM visualizations confirm that the CNN focuses on semantically meaningful regions—such as vegetation, water bodies, and property boundaries—particularly for high-priced homes. This validates that the multimodal model leverages visual features in an interpretable and financially relevant manner rather than learning spurious patterns.

Key Takeaways

- Integrating satellite imagery leads to **substantial accuracy improvements** in property price prediction.
- Early feature fusion enables effective combination of tabular and visual representations.
- Grad-CAM analysis enhances trust by confirming that the model attends to meaningful environmental features.